Evaluation of HRRR-ARW Forecasts

Summary Report

MD&E Task 09.5.24E1

Juanzhen Sun, David Dowell, Mei Xu, Ying Zhang, James Pinto, and Matthias Steiner

National Center for Atmospheric Research (NCAR)
Boulder, Colorado

19 September 2009

1. Purpose of the Study

The work reported in this document is the result of a collaborative effort between the NCAR Research Applications Laboratory (RAL)—i.e., both the Model Development and Enhancement (MD&E) and Convective Weather (CW) teams—and the NOAA Earth System Research Laboratory (ESRL) Global System Division (GSD). The purpose of the work reported here is to evaluate the performance of the High-Resolution Rapid Refresh (HRRR) 3-km WRF-ARW forecasts of convective events through case studies and a real-time qualitative assessment. The case studies were conducted to evaluate the sensitivities of HRRR forecasts with respect to the initialization method, grid resolution, microphysics scheme, domain size, and lateral boundary effects. The real-time assessment focused on the overall HRRR performance in terms of forecast success and failure to predict convective weather. Some of the failure cases were rerun to diagnose the sources of error that may have contributed to missing the forecast.

2. Real-time Qualitative Assessment of HRRR Performance

2.1 CoSPA Summer 2009 convection forecast exercise

The HRRR model forecasts were reviewed regularly during the summer 2009 as part of the Collaborative Storm Prediction for Aviation (CoSPA) forecast system monitoring. The NCAR Scientists on Duty (SoD) met weekly with other CoSPA program participants to discuss availability and performance of the HRRR model during the previous week. From these discussions, two main qualitative impressions emerged for the summer 2009 HRRR forecasts:

(1) The HRRR performed particularly well during the late afternoon and into nighttime, when the model often produced convective systems with realistic structure, and with relatively small timing and location errors.

(2) The HRRR performance was not as good in the morning and early afternoon. Often, the model did not produce new convection as readily as the atmosphere during this time period. Also, the model often struggled to maintain convective systems that were ongoing at the model-initialization time, and moving into regions where solar heating was occurring.

An example of a very good late afternoon forecast by the HRRR is shown in Fig. 1 for 20 August 2009. In this case, line segments began forming at approximately 1800 UTC and by 2300 UTC (upper-left panel in Fig. 1) a relatively solid line had formed; this line was significant enough to affect air traffic. All available HRRR forecasts valid at 2300 UTC indicated SSW-NNE line orientations, consistent with the observed line orientation. However, differences among the forecasts were apparent in line location and storm coverage along the line. Based on a subjective assessments by the SoD's, one could speculate that forecasts valid during the late afternoon and into nighttime are typically relatively good for one or more of the following reasons:

(a) observations used to initialize these forecasts are relatively plentiful, because more observations tend to be available during the day (e.g., aircraft observations) than at night,

(b) the information provided by surface observations is greater during the day, owing to a higher correlation with the planetary boundary layer (PBL) state through a deep layer, and

(c) nighttime convective systems are generally larger than daytime systems and are thus more predictable.

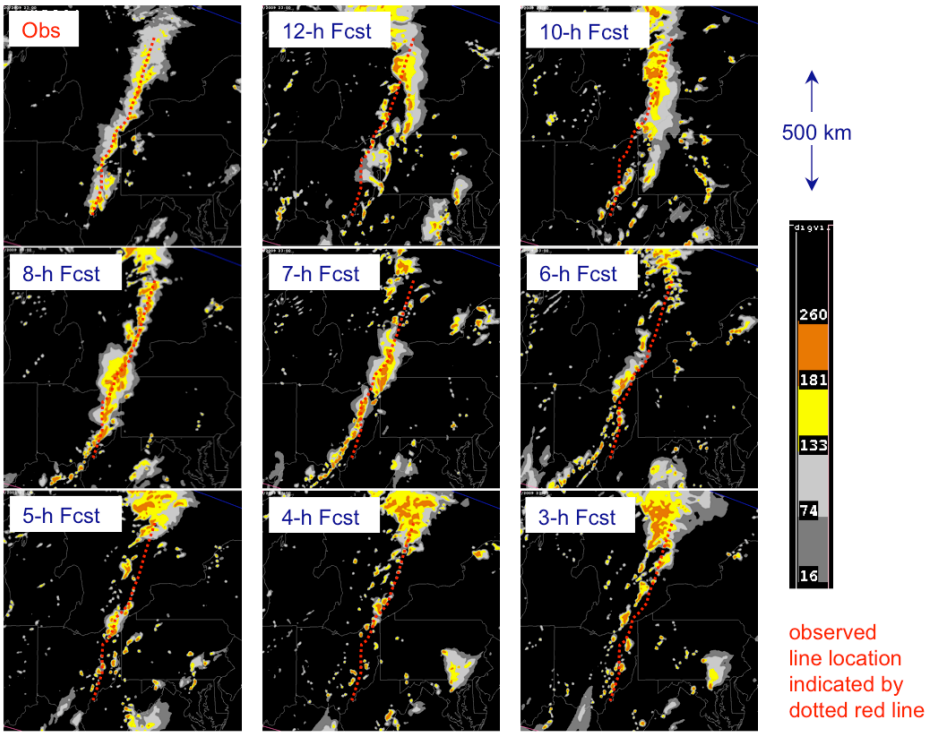
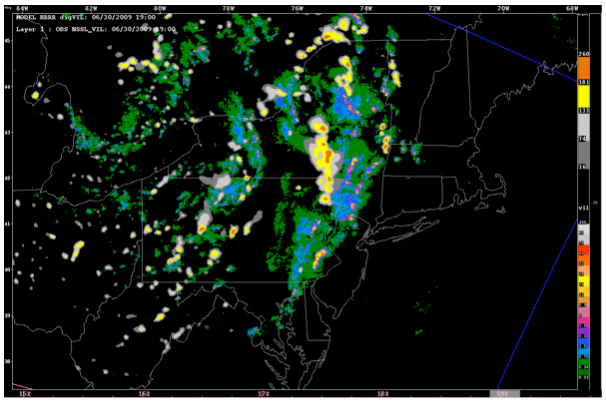
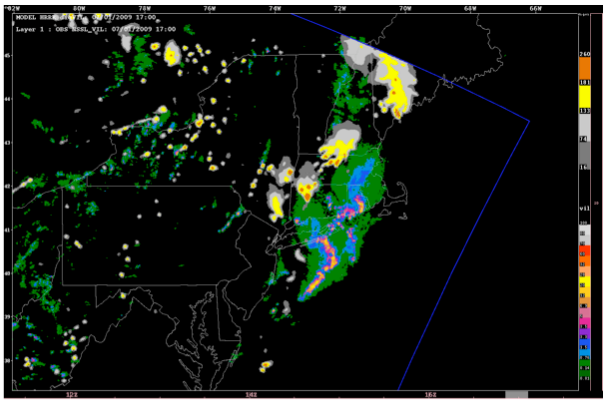


Fig. 1. Observed (upper-left panel) and HRRR (all other panels) digital VIL for a convective line in the Midwest at 2300 UTC 20 August 2009.

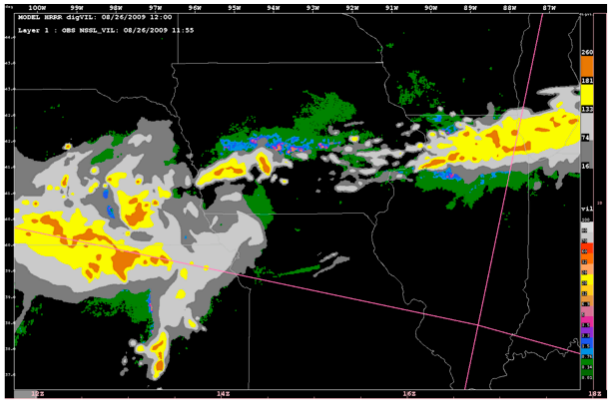
In contrast, three morning/early afternoon cases for which the HRRR underpredicted the coverage of convective storms are shown in Fig. 2. For the 30 June 2009 case (Fig. 2a), the model failed to predict the eastward extent and widespread coverage of convection from southeast PA to western MA, moving toward the Northeast Corridor airports. The following day (Fig. 2b), the model failed to predict a convective system that began as a small area of convective storms over NJ around 1300-1400 UTC (not shown) and then expanded as it moved toward the northeast and joined with other convective systems from Long Island to CT. Thirdly, for the 26 August 2009 case (Fig. 2c, d), the HRRR initially was successful in forecasting convection associated with a mesoscale convective vortex and a W-E oriented front to its east; however, later the convection dissipated in the model while the observed convection persisted as it moved eastward into southern IA.



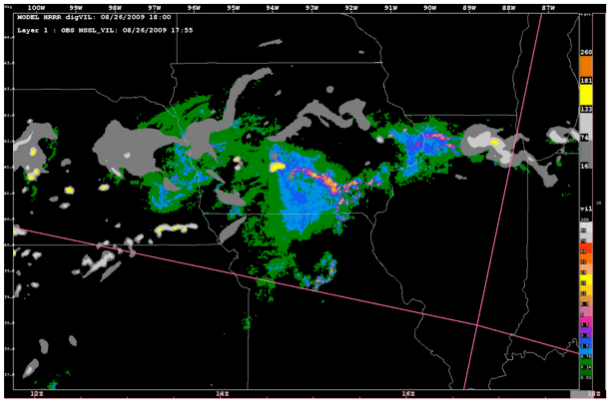
(a) 1900 UTC 30 June 2009



(b) 1700 UTC 1 July 2009



(c) 1200 UTC 26 August 2009



(d) 1800 UTC 26 August 2009

Fig. 2. HRRR digital VIL (gray/white/yellow/orange shading) overlaid on observed VIL (green/blue/purple/yellow/red shading). (a) 1900 UTC 30 June 2009 (9-h HRRR forecast). (b) 1700 UTC 1 July 2009 (8-h HRRR forecast). (c) 1200 UTC 26 August 2009 (4-h HRRR forecast). (d) 1800 UTC 26 August 2009 (10-h HRRR forecast).

The challenges of predicting morning convective storms / systems in summer 2009 motivated further analysis of the cases shown in Fig. 2. Further examination of these cases is described later in the "Error Diagnoses and Forecast-Sensitivity Studies" section.

2.2 VORTEX2 convective forecast exercise

The Verification of the Origins of Rotation in Tornadoes Experiment 2 (VORTEX2) is a multi-agency program to study tornadoes, tornadic storms (supercell thunderstorms), and numerical weather prediction of severe storms. The first field phase of VORTEX2 occurred 10 May – 13 June 2009. D. Dowell's involvement in both CoSPA and VORTEX2 provided a unique "bonus" opportunity to assess HRRR performance. Storm forecasting and targeting was based on consensus of the principal investigators (PIs), and approximately a dozen VORTEX2 PIs were regularly looking at HRRR forecasts.

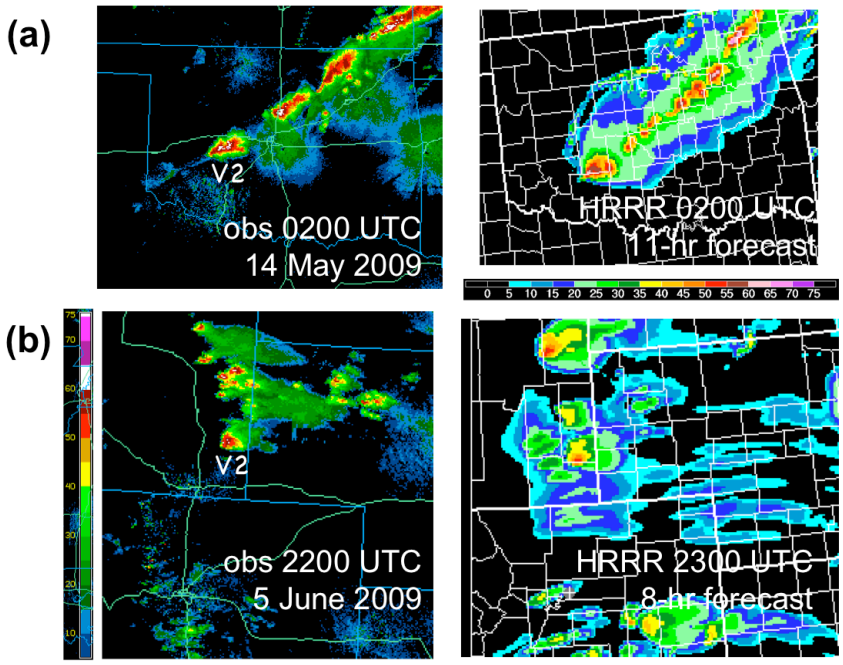


Fig. 3. Observed (left, dBZ) and HRRR composite reflectivity (right, dBZ) for two cases during VORTEX2 (images from College of DuPage and NOAA ESRL). The "V2" symbol indicates storms VORTEX2 targeted for data collection. In both cases, the HRRR indicates a convective storm with significant rotation (not shown) within 50 km and 1 h of the observed supercell thunderstorm. (a) 13 – 14 May 2009 supercell in western Oklahoma. (b) 5 June 2009 tornadic supercell in southeast Wyoming.

The domain for VORTEX2 field operations covered the central United States plains, and the project has been studying isolated severe-storm events during the afternoon and evening. Thus, impressions of HRRR performance are specific to this region, time of day, and storm type. Overall, the impressions of HRRR performance during VORTEX2 in 2009 are summarized as follows:

(1) The HRRR had a very high probability of detection for severe storms. In fact, for all days on which VORTEX2 intercepted supercell thunderstorms (9 days), the HRRR indicated strong convective storms within ~100 km of the region where VORTEX2 collected observations. Two examples of HRRR forecasts during VORTEX2 operations are shown in Fig. 3. In addition, on numerous occasions, when the HRRR indicated no significant convective storms in a candidate target region that VORTEX2 was considering, the atmosphere indeed failed to produce significant convective storms in this region.

(2) False alarms occurred in the HRRR forecasts occasionally, and these were generally of two varieties: (a) the model produced strong convective storms in a particular region, but the atmosphere did not; and (b) the model produced too many convective storms in a particular region, resulting in upscale transition to a larger-scale, incorrect storm mode.

(3) Timings of HRRR-forecast and observed storms were typically within 1-2 hours of each other for storms that VORTEX2 targeted.

(4) The HRRR provided very reliable forecasts of storm motions, which were helpful for planning deployments of mobile instruments.

3. Error Diagnoses for Selected Cases

As described previously, SoD's noted reduced model performance for morning and early afternoon HRRR convective forecasts during the summer 2009 CoSPA forecast assessments. These impressions motivated further examination of the cases shown in Fig. 2a and b, for which the model underpredicted the areal coverage of convection in locations important for air traffic. Specifically, HRRR forecasts of surface fields and clouds were examined to determine how well the model handled these aspects of the mesoscale situation.

To reduce computational requirements for forecast-sensitivity experiments, the HRRR forecasts were rerun for relatively small subdomains (300 × 300 gridpoints, whereas the operational HRRR in summer 2009 included 1000 × 700 gridpoints). Convective forecasts in the large (operational) and small (rerun) domains were very similar (not shown), and thus the reruns in small domains were considered representative for further analysis.

The lowest grid level in the HRRR is at approximately 8 m AGL, whereas fields at the standard 2-m surface-observation screen height are diagnosed from the WRF prognostic fields. During the comparison of HRRR fields to surface observations, one issue was quickly discovered: A *large discrepancy develops between the dewpoint / water vapor at the lowest model level (8 m) and that diagnosed at 2 m during the day* (Fig. 4). Model profiles (left side of Fig. 4) suggest that the dewpoint differences between 2 and 8 m AGL are too large to be physically plausible. Large (superadiabatic) differences also develop in temperature, but such differences could be physically plausible in sunny locations when the sun is overhead, particularly in summer. The discrepancy between daytime dewpoint temperatures at 2 and 8 m AGL was also seen in the model on other occasions, and for locations both in the eastern US and in the plains (not shown). This issue, which could indicate an error within the RUC land-surface model in WRF and/or an error in the interface between the RUC land-surface model and other WRF modules, has been reported to NOAA/ESRL/GSD.

For the remainder of the discussion, the comparison between model and surface observations is for temperature at the lowest grid level (8 m AGL) in the model. HRRR forecasts have been rerun for several cases, with two cases shown in Figs. 5 and 6, corresponding to cases for which areal coverage of convection was underpredicted (Figs. 2a and 2b). These forecasts were rerun both with the standard HRRR configuration, which uses the RUC land-surface model, and with a configuration that is otherwise identical but uses the NOAH land-surface model. In the 30 June 2009 case, a comparison of the visible satellite image and the model downward radiation flux at the surface (related to cloud cover, and shaded in a manner to allow qualitative comparison to the satellite image) shows that the model correctly predicted clear skies from VA to

NJ to CT (red circled region in Fig. 5). In the sunny regions, the model warms too slowly at low levels, as indicated by the abundance of green/blue dots in the circled region.

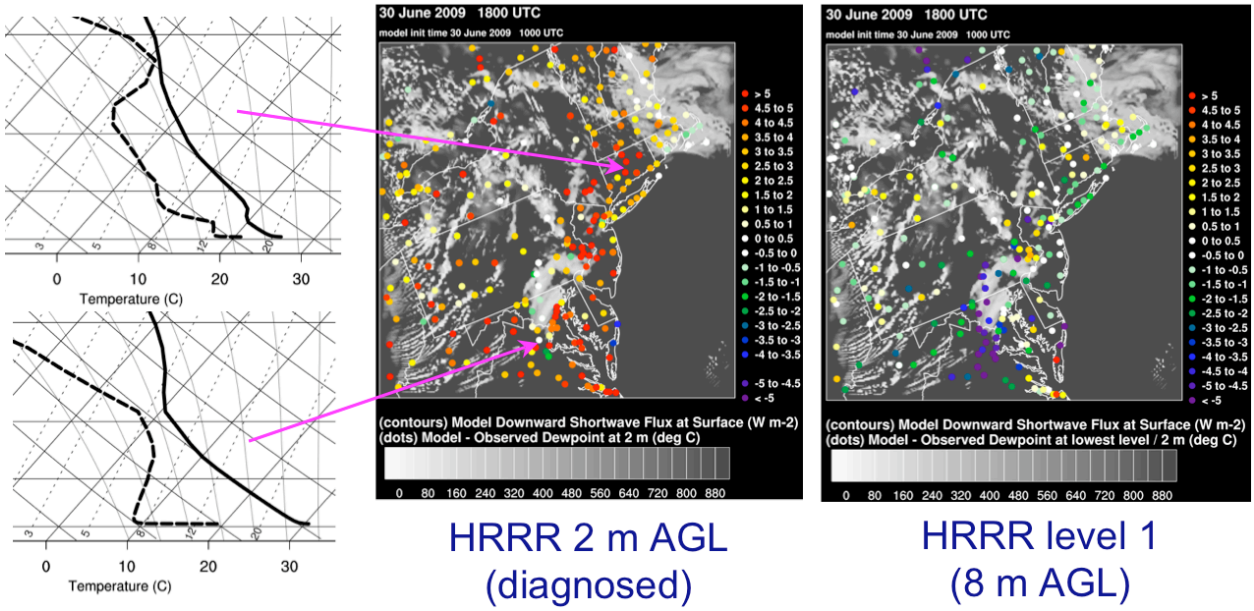


Fig. 4. Differences between HRRR (rerun initialized at 1000 UTC 30 June 2009) and observed 2-m AGL dewpoint temperatures (dots, color coded by difference in °C) and downward shortwave radiation flux at the surface (shading, in W m^{-2}) at 1800 UTC 30 June 2009. Differences are computed for the fields at the lowest model level (approximately 8 m AGL; right) and for the fields that are diagnosed by the WRF model at 2 m AGL (left). Surface observations were obtained from NOAA's Meteorological Assimilation Data Ingest System (MADIS). Skew *T* – log *p* model profiles [temperature (dewpoint) indicated by solid (dashed) lines] are shown on the left, for two grid points at the indicated locations.

A similar trend is seen in the 1 July 2009 case (Fig. 6). In the sunny region from eastern VA to eastern PA, the model is again too slow to warm the low levels, and there is an abundance of green/blue dots (negative values of model minus forecast) in the southernmost circled region. For the two cases, the underforecast surface temperature in sunny regions appeared around 1200-1300 UTC and persisted into the afternoon.

In both the 30 June and 1 July cases, the magnitude of the difference between model and observations is less for the NOAH than for the RUC land-surface model. *Thus, there appears to be some potential for improving the HRRR surface-temperature forecasts by tuning the RUC land-surface model and/or using the NOAH land-surface model.* More cases must be examined to determine if the forecast differences related to the land-surface models are systematic and/or seasonal.

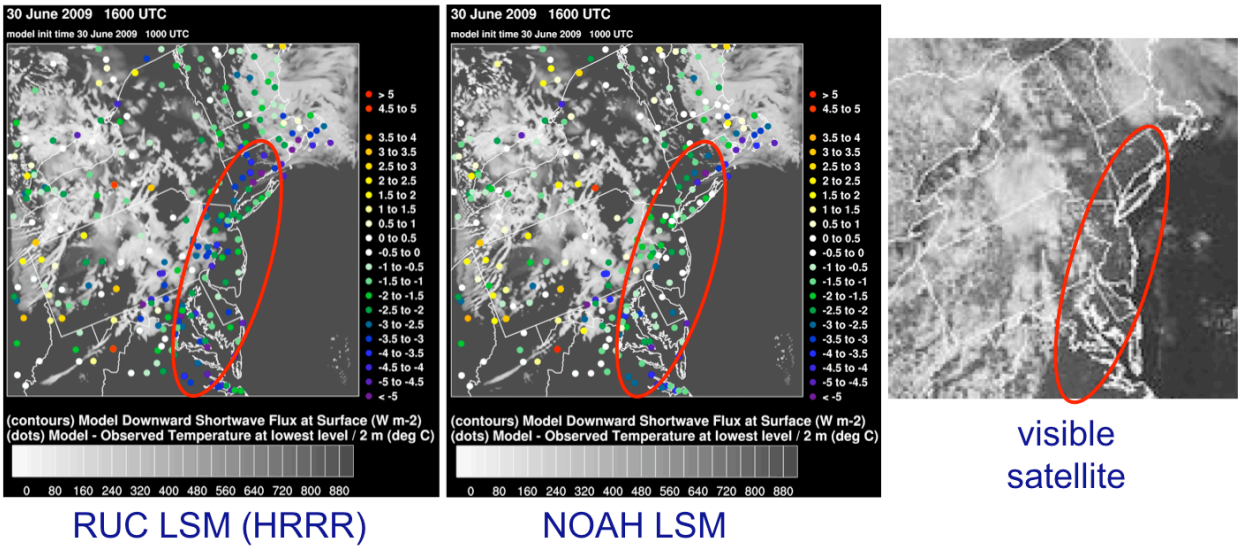


Fig. 5. Differences between HRRR (rerun initialized at 1000 UTC 30 June 2009) 8-m AGL and observed 2-m AGL temperatures (dots, color coded by difference in °C) and downward shortwave radiation flux at the surface (shading, in W m^{-2}) at 1600 UTC on 30 June 2009. The model was run with the default RUC land-surface model (left panel) and the NOAH land-surface model (middle panel). A visible satellite image at 1630 UTC is also shown (right panel).

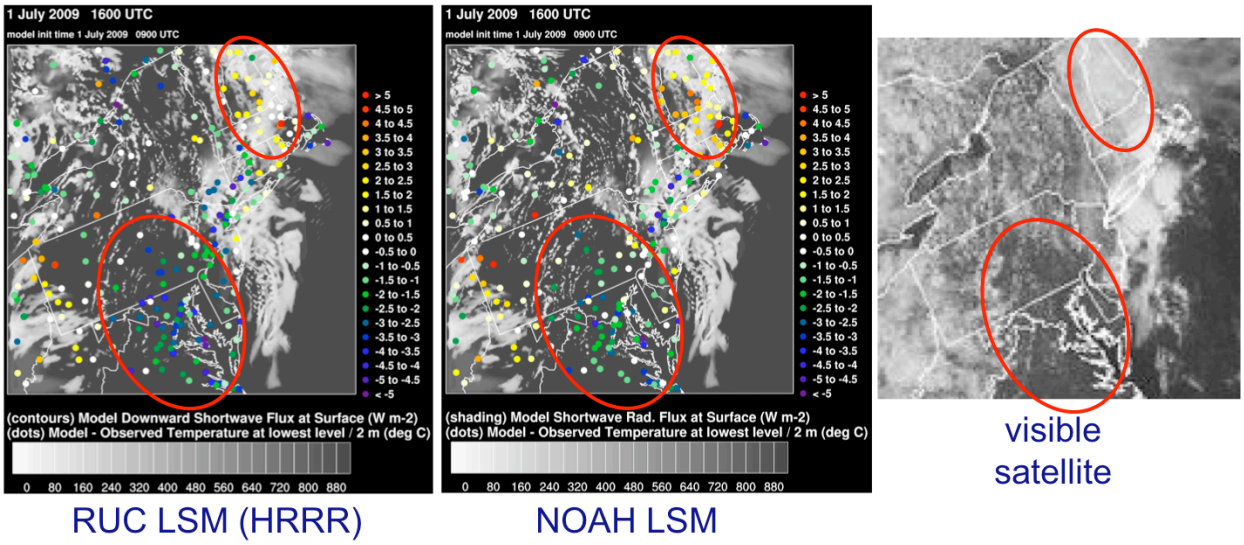


Fig. 6. As in Fig. 5, except the HRRR reruns were initialized at 0900 UTC on 1 July 2009, and the results are shown for 1600 UTC on 1 July 2009.

Also of note in Fig. 6 is the circled region over New England, where the model correctly indicates clouds, but where the low levels of the model are too warm. The net result—with sunny areas being too cool and cloudy areas too warm—is that *the daytime low-level temperature contrasts between sunny and cloudy regions were typically less in the*

model than in the atmosphere for the summer 2009 cases examined so far. Analysis is ongoing to determine the generality of these results, and the implications for convective forecasts. Results of this ongoing analysis will be described in reports at a later date.

Table 1. List of the events or episodes examined in case studies and sensitivity test.

Case date	*Forecast cycles*	*Phenomena of interest*	*Tests conducted*
20080727	2712 2715	Line convection in the east, the initiation, organization, movement, and dissipation	Initialization sensitivity Grid increment sensitivity
20080805	0418 0421 0500 0503 0506	MCSs, multi-cell clusters' formation, development, dissipation and re-initiation	Initialization sensitivity Grid increment sensitivity Microphysics test Lateral boundary test Radar data assimilation test
20080815	1512 1515	Widespread convection especially in the N.E.	Initialization sensitivity Grid increment sensitivity
20090513	1312-1412 Every 3 h	Line storms formed along OK-NE-MI and moved eastward	Rerun with CONUS domain
20090514	1409	Widespread convection along southeastern and Gulf coasts	Lateral boundary test
20090515	1512 1515	Line of storms formed in GA - SC coast and propagated westward	Initialization sensitivity Grid increment sensitivity Lateral boundary test
20090611	1100 1106	Super-cell or multi-cell clusters in northern Texas	Lateral boundary test
20090706	0612-0712 Every 3 h	Widespread convection along the Gulf coast	Rerun with CONUS domain
20090716	1612-1712 Every 3 h	Widespread convection along the Gulf coast	Rerun with CONUS domain

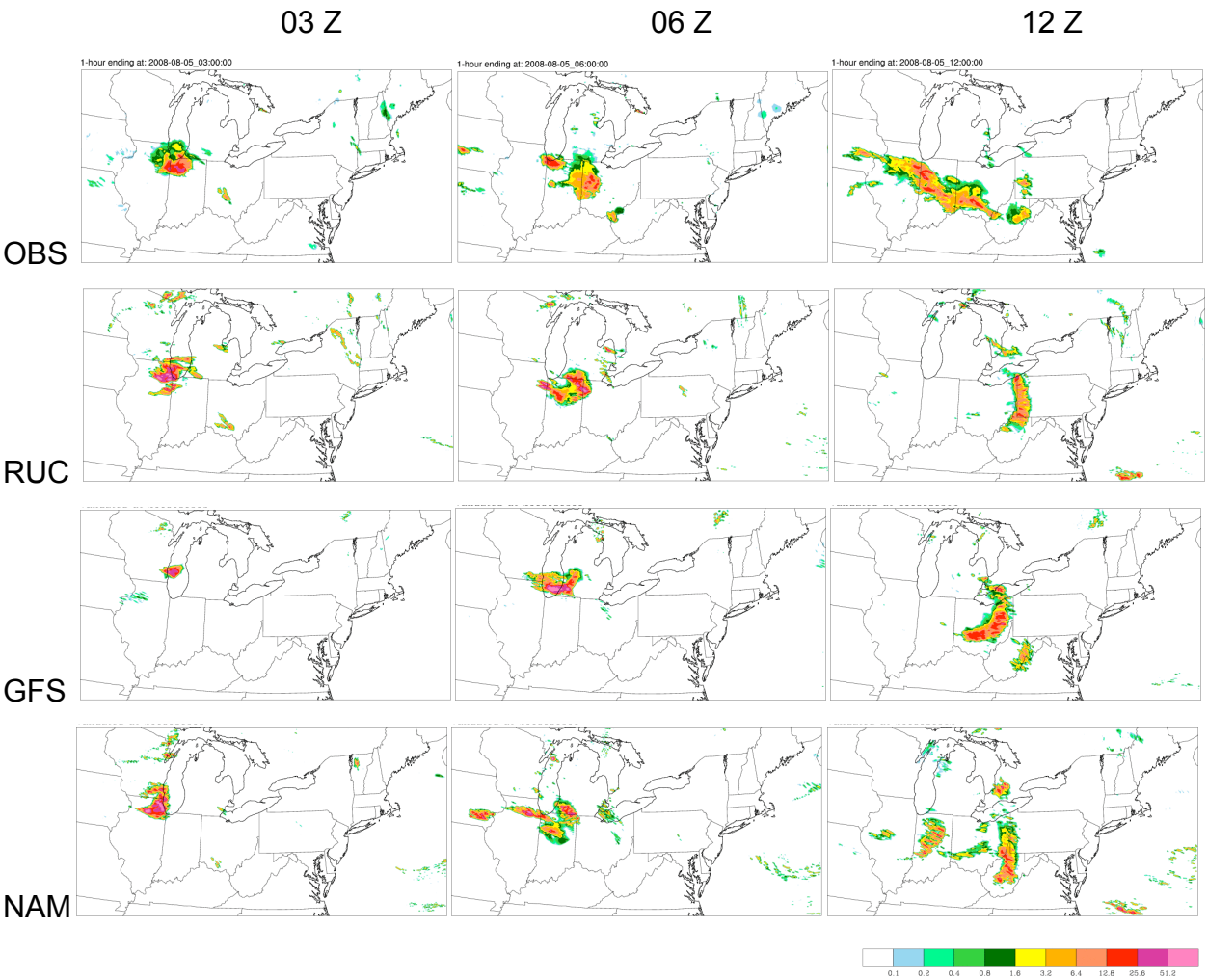


Fig. 7. One-hour rainfall rate from Stage IV (OBS), and model forecasts initialized with RUC (RUC), 0.5° GFS (GFS), and 40 km NAM (NAM), respectively. The model initialization time is 00 Z of 20080805 and forecast validation time are 03, 06 and 12 Z.

4. HRRR Evaluation by Sensitivity Tests of Selected Convective Cases

Forecasts from the HRRR have been evaluated extensively through case studies and sensitivity tests. Interesting cases that span a variety of weather regimes have been identified from this summer (2009) as well as the previous summer (2008). Table 1 summarizes the case studies and sensitivity runs that are described in this section. The factors examined are model initialization method, grid increment, microphysics scheme, domain size and proximity of domain boundaries. Some of the cases were selected for sensitivity testing, especially, because the HRRR forecasts failed in certain aspects. Therefore the skill scores discussed in this chapter may not be fully representative of the overall skills of HRRR for the entire season. Instead, particular performance problems in specific conditions are reviewed here. Moreover, one has to keep in mind also that the HRRR evolved from 2008 to 2009.

4.1 Forecast sensitivity with respect to model initialization

The operational HRRR forecasts are initialized with 13 km RUC datasets that are enhanced with radar reflectivity data assimilation. Therefore, the HRRR forecasts may be considered as "hot-started" and enhanced for very short-term forecasts. Visual observations show that HRRR forecasts still experience a spin-up period, but it is relatively short (30 minutes to an hour). General observations also reveal that HRRR forecasts, valid at the same time but with different lead-times, differ significantly. The performance skills usually peak in the 1-2 h forecasts and then decrease with forecast length. A suite of sensitivity runs are conducted using the same model configurations as HRRR, but initializations with 40 km NAM and 0.5° GFS datasets. The purpose for these runs is to characterize the benefit of a radar-enhanced RUC initialization in HRRR and find out how long the benefit lasts as the forecast length increases.

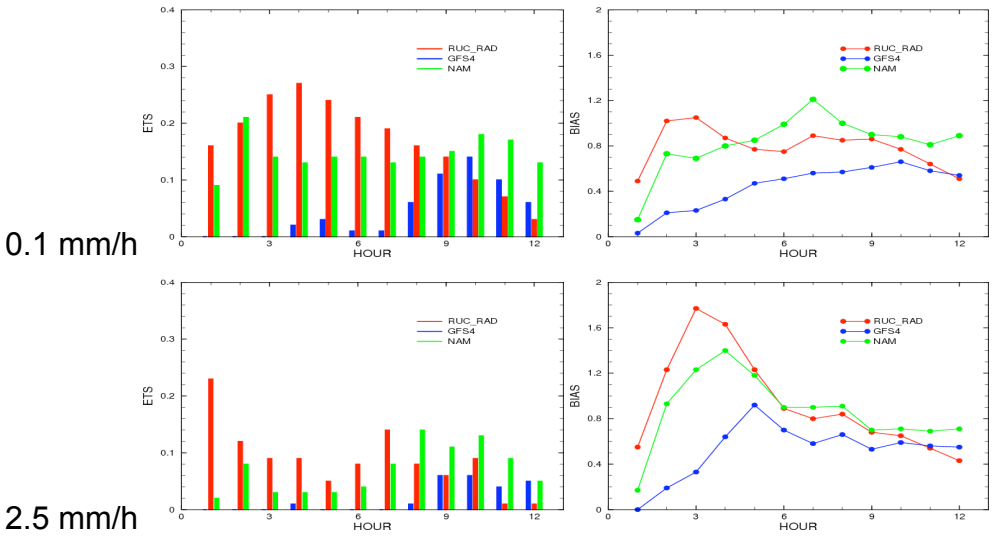


Fig. 8. Equitable Threat Score (left) and Bias score (right) of the forecasts with three different initializations: RUC (red), 0.5° GFS (blue), and 40 km NAM (green), for case 2008080500.

Forecasts with the three different initializations are compared and evaluated both subjectively and objectively. For example, snapshots of hourly rainfall rate from the three forecasts and Stage IV observations for case 2008080500 are shown in Fig. 7, while Fig. 8 shows conventional skill scores for the same forecasts. In this case, an intensive convective system already existed at the forecast initialization time and subsequently developed into a Mesoscale Convective System (MCS). As new storms continued to initiate throughout the night, they joined the old one to form a large region of storm activities. All forecasts predicted the MCS. However, vast differences are seen in these forecasts from different initial conditions in terms of storm locations, trajectories and life cycles. In the first hour, forecasts with NAM and GFS significantly underpredict convection due to a slower spin-up from the cold start (not shown). At 6 h, the RUC initialized HRRR forecast still had some advantage over the other two

forecasts. By 12 h, all forecasts had significantly departed from the observations. In terms of objective skill scores, the RUC initialized forecast clearly outperformed in the 1-7 h forecast range, while the NAM and GFS initialized forecasts caught up in the 8-12 h range. For this case, the NAM initialized forecast also did pretty well in forecasting light precipitation.

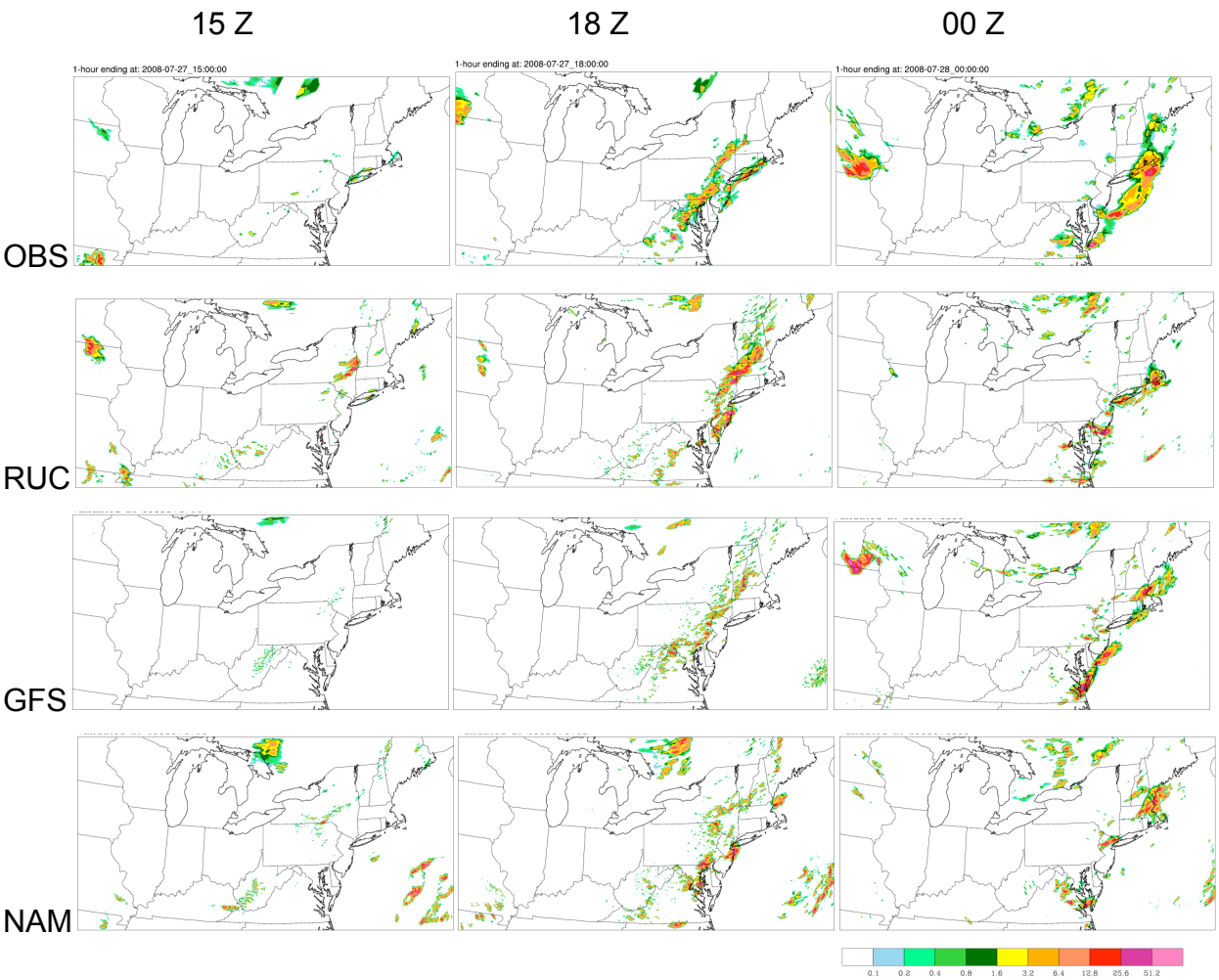


Fig. 9. One-hour rainfall rate from Stage IV (OBS), and model forecasts initialized respectively with RUC (RUC), 0.5° GFS (GFS), and 40 km NAM (NAM). The model initialization time is 12 Z of 20080727 and forecast validation time are 15, 18 and 00 Z of 0728.

Precipitation rates and skill scores for a case of line storms on 20080727 are depicted in Figs. 9 and 10. The line storms initiated around 15 Z, 3 hours after the forecast initialization at 12 Z. None of the forecasts captured the initial storms well in their 3-5 h forecasts. Nevertheless, the 6 - 9 h forecast of RUC initialized HRRR better predicted the storms in their mature stage.

The model skills vary significantly from case to case. Skill scores for another case, initialized at 12 Z of 20080815, are shown in Fig. 11. The RUC initialization (i.e., HRRR) performed significantly better for 1-5 h before losing skills. Fig. 12 shows the forecast skills averaged over six forecasts. On average, the skill curves, especially those for heavy precipitation, place HRRR as the favorable forecasts for the 0 - 9 h range. The 1-6 h forecasts from HRRR are significantly better than the other two models, while the benefit gradually levels off in later hours. The individual cases show that for the 7-10 h range, the results are mixed and dependent upon the convective regime. There also seems to be a tendency in HRRR to overforecast in the 3-5 h range, shortly after spin-up.

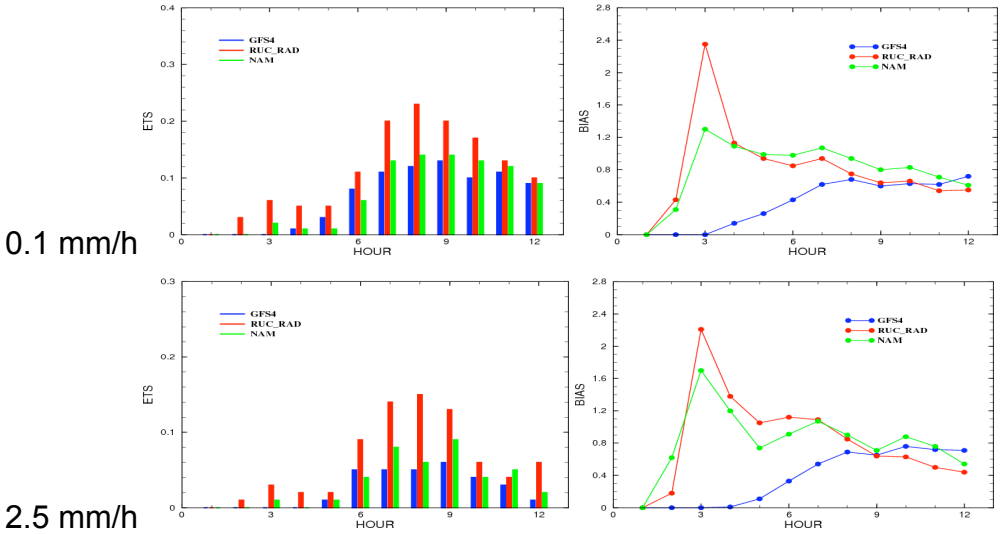


Fig. 10. Equitable Threat Score (left) and Bias score (right) of the forecasts with three different initializations: RUC (red), 0.5° GFS (blue), and 40 km NAM (green), for case 2008072712.

4.2 Forecast sensitivity with respect to grid resolution

It is generally recognized that numerical models with horizontal grid resolution of 1 - 5 km could resolve atmospheric convection explicitly. However, the exact cut-off is not well established and is still a topic of research. In 2008 the HRRR used a grid increment of 3 km, while 3.1 km was utilized in 2009 (for reasons of domain setting and HRRR run time). How good are these model configurations in resolving the convective storms of aviation interests? Given the expected increase in computer power in the near future, should we pursue a further reduction in model grid increment or larger domain coverage? For example, a 27-fold increase in computer resources will be required to run the model at a horizontal resolution of 1 km instead of 3 km. A set of sensitivity runs have been conducted to gain some preliminary understanding of the effect of model resolution, as initial guidance toward further HRRR model development.

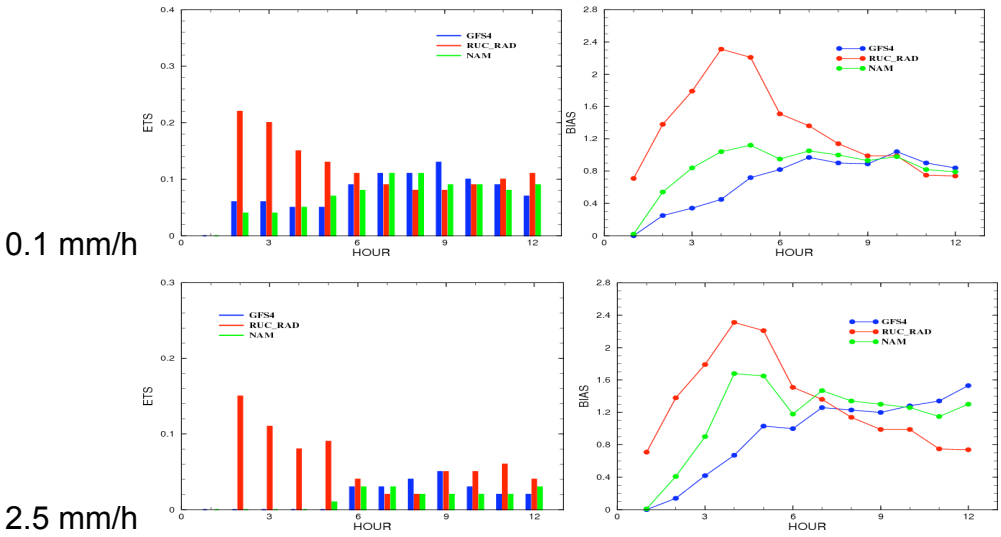


Fig.11. Equitable Threat Score (left) and Bias score (right) of the forecasts with three different initializations: RUC (red), 0.5° GFS (blue), and 40 km NAM (green), for case 2008081512.

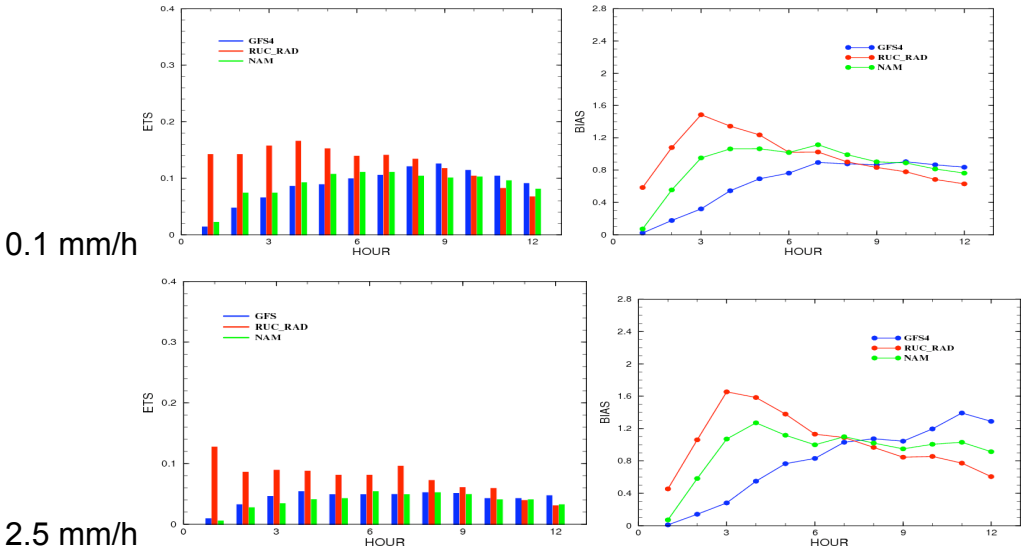


Fig. 12. Equitable Threat Score (left) and Bias score (right) of the forecasts with three different initializations: RUC (red), 0.5° GFS (blue), and 40 km NAM (green), averaged over six forecasts.

In addition to 3 km HRRR forecasts, model runs with 1 km and/or 5 km grids have been conducted for several cases within different convection regimes (Table 1). These model runs were all initialized with RUC datasets, and they differ only in grid resolution. The 1 km and 3 km forecasts for 2008080500 (the nocturnal MCS case) are shown in Fig. 13. It is found that the differences between the 1 km and 3 km forecasts are subtle and mostly in the small-scale details. The major problems with the 3 km forecasts were not alleviated by the use of a 1 km grid. The differences between the 5 km and 3 km

forecasts are more significant (not shown): The movement of the predicted convective system is somewhat slower when 5 km grid is used. In summary, it didn't make a great difference whether 3 km or 1 km model grids were used in this case.

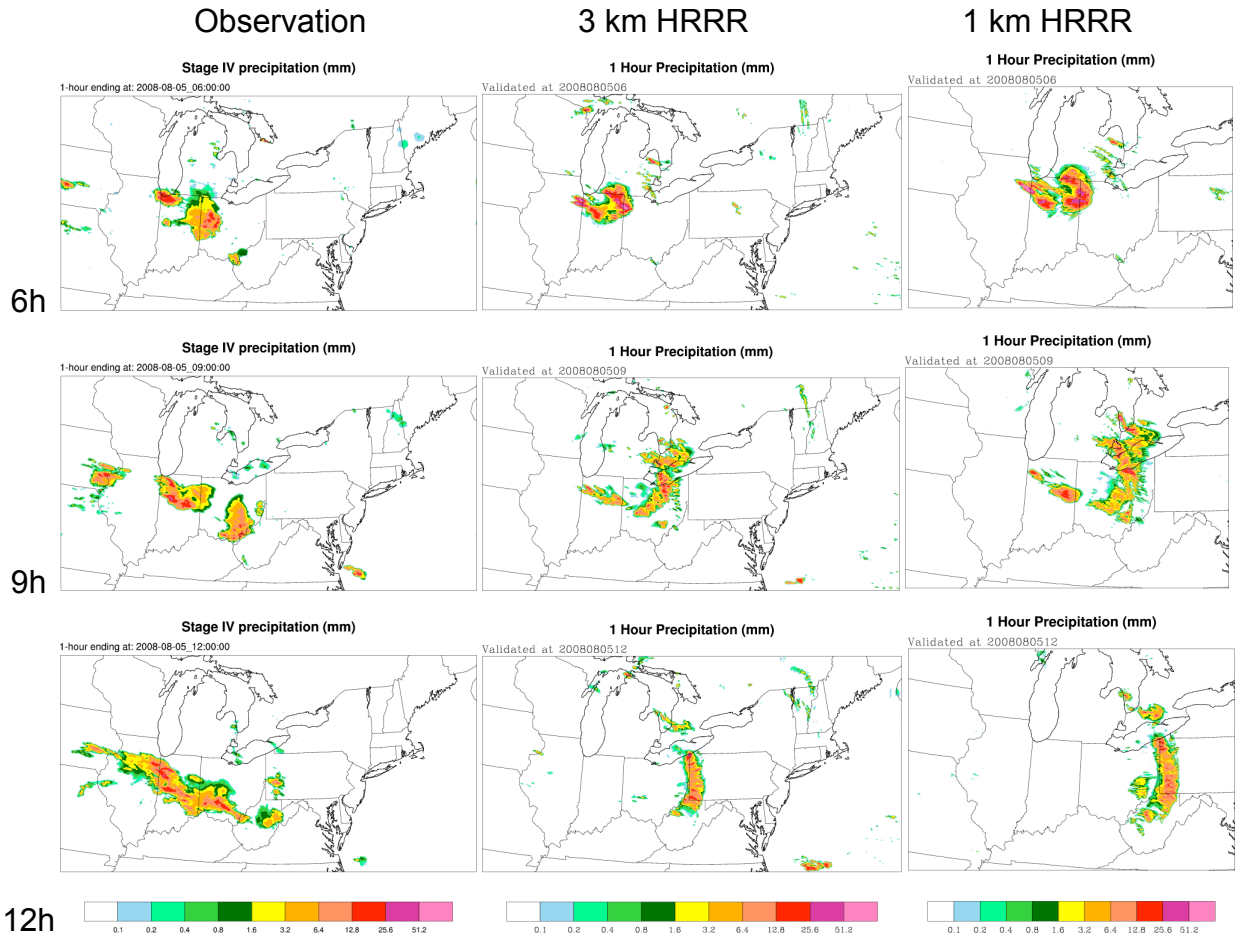


Fig. 13. One-hour rainfall rate from Stage IV and model forecasts using 3 km and 1 km grids. The models are initialized at 20080805 00 Z and 6 h, 9 h and 12 h forecasts valid at 06, 09 and 12 Z of 20080805 respectively are shown.

Fig. 14 shows the grid increment sensitivity in another forecast, initialized at 12 Z of 20080727. For this case of line storm initiation and evolution, again there seems to be no major difference between the 1 km and 3 km runs at shorter lead-time (3 h - 9 h). However, at 12 h, the two forecasts seem to have considerable differences. The 1 km forecast placed the storm in the middle part of the line at a more advanced location.

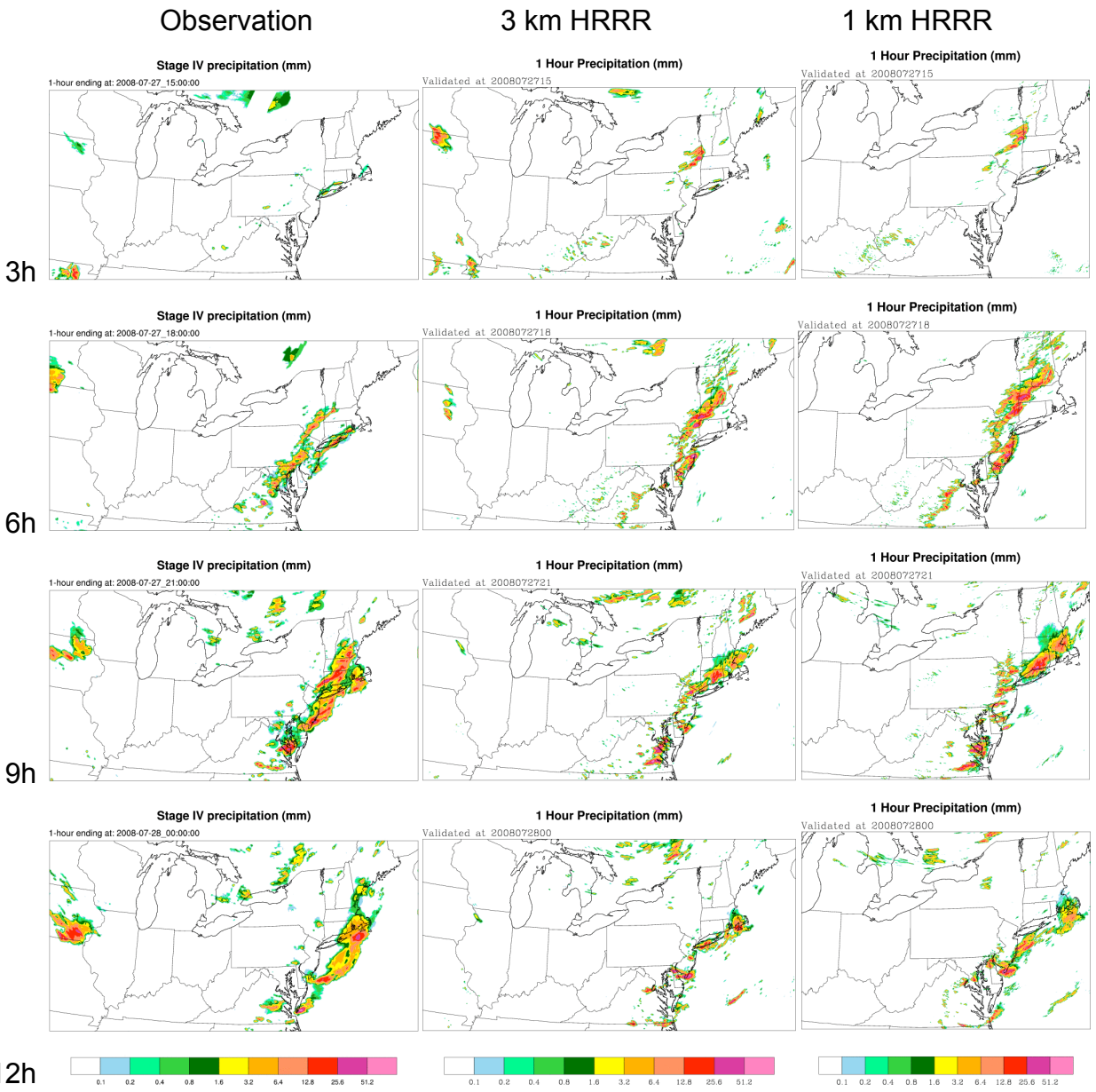


Fig. 14. One-hour rainfall rate from Stage IV and model forecasts using 3 km and 1 km grids. The models are initialized at 20080727 12 Z and 3h, 6 h, 9 h and 12 h forecasts valid at 15, 18, 21 and 00 Z of 20080728 respectively are shown.

4.3 Performance of the new Thompson microphysics scheme

The numerical model core of HRRR in 2008 and 2009 is based on WRFv3.0 that was released in the spring of 2008. The one-moment Thompson microphysics scheme was used. As of spring 2009, a two-moment Thompson microphysics scheme had become available and was released with WRFv3.1. Offline tests were conducted for 5 forecast cycles to find out the effect of this new microphysics scheme on HRRR forecasts. Of particular interests are the CPU time required to run the new microphysics scheme and

the effect of the new scheme on the prediction of storm structure, evolution and precipitation amount. Results show that with the new Thompson scheme each 12 h forecast takes about 3.5% – 3.8% more CPU time than the forecast with the old Thompson scheme (in current HRRR).

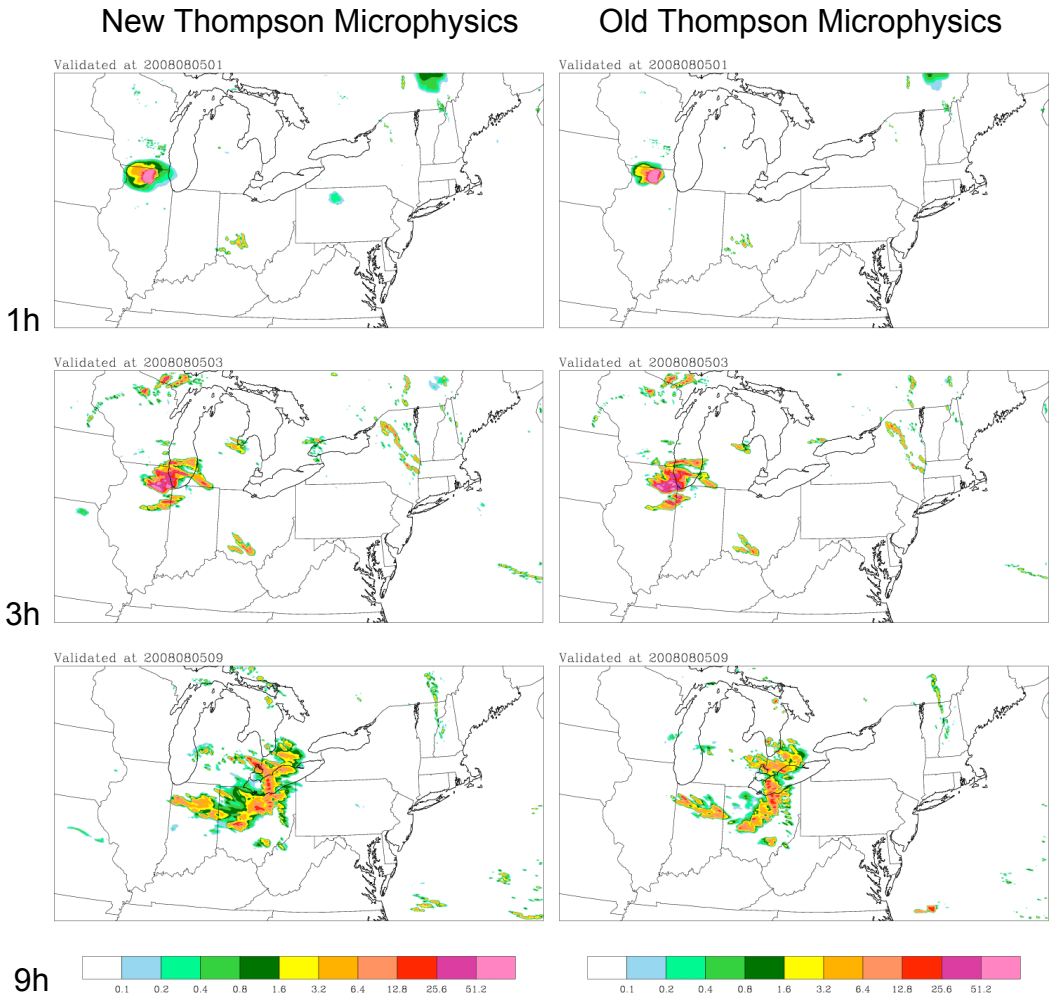


Fig. 15. Hourly precipitation rate (mm/h) from model forecasts using the new (left) and old (right) Thompson microphysics scheme. The forecasts are initialized at 20080805 00 Z and the validation time are 01, 03, and 09 Z of 20080805.

Figs. 15 and 16 compare the hourly precipitation forecasts from the new scheme with those from the old scheme, for selected lead-time in two forecast cycles. The most evident difference is that the new scheme results in more extensive light precipitation, especially in areas behind the leading edges of the MCSs. There are also some differences in terms of the details of heavy precipitation cores. For most forecast hours, there is no significant impact on the storm evolution or movement. However, for the forecast cycle starting at 2008080421 (Fig. 16), the new microphysics does make some

visible differences in the overall storm structure and locations, presumably due to feedbacks of the cold pool.

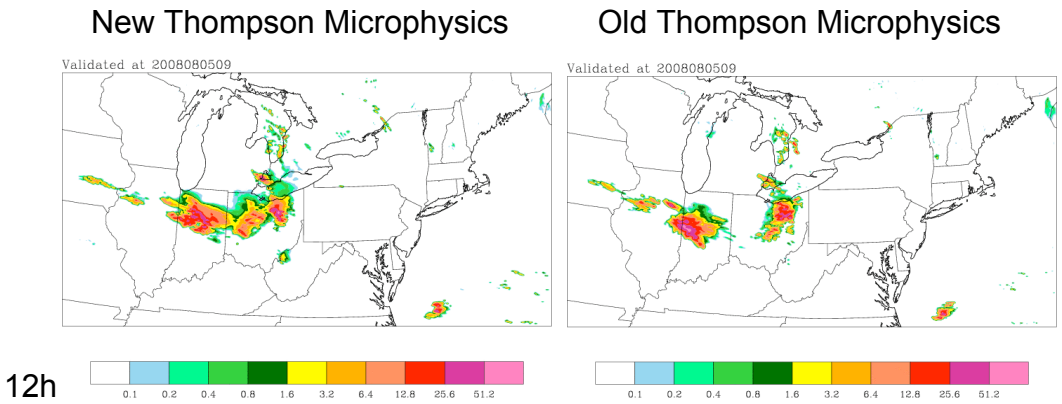


Fig. 16. Same as Fig. 15 but for 12 h forecasts starting from 21 Z of 20080804. The validation time is 09 Z of 20080805.

In terms of objective skills (BIAS, CSI and ETS) (not shown), a significant improvement is seen in the first hour after initialization: The new scheme seems to have a faster spin-up. Beyond the first hour, the changes in skills are mixed and mostly (but not always) insignificant. In general, the new microphysics scheme increases the precipitation area and, therefore, in most cases increases the skills of previously underpredicted situations and decreases the skills where it previously overpredicted precipitation.

4.4 Sensitivity of HRRR forecast with respect to proximity of domain boundaries

According to real-time monitoring SoD reports, the HRRR often failed to forecast scattered convection along the southeast U.S. (Atlantic and Gulf of Mexico) coasts. This led us to carry out several lateral boundary sensitivity runs with an extended southeastern model domain boundary (Table 1). Sensitivity tests for lateral boundary proximity have also been conducted by moving away the boundary upstream from other major convective events of interest. In addition, forecasts for a CONUS domain have been conducted to investigate how the new configuration may affect such forecasts. These CONUS runs will be utilized also for CoSPA blending development. A summary of two case studies, for events that occurred on 15 May and 11 June 2009, respectively, is given here.

The 15 May 2009 case exhibits scattered convection in the southeast (Fig. 17). Around 11 – 12 Z on that day, some echoes started to pop up along the Georgia and South Carolina coasts. As more convective cells started to appear, they initiated further inland and gradually formed into a line parallel to the coast. From 18 – 20 Z, the storms were well organized into a thin line of convection pushing inland, getting as far as 100-150 km away from the coast. By 22 Z, they merged with the scattered storms further northwest over the Appalachians behind a warm front to form wide spread scattered convection over the entire southeastern U.S.

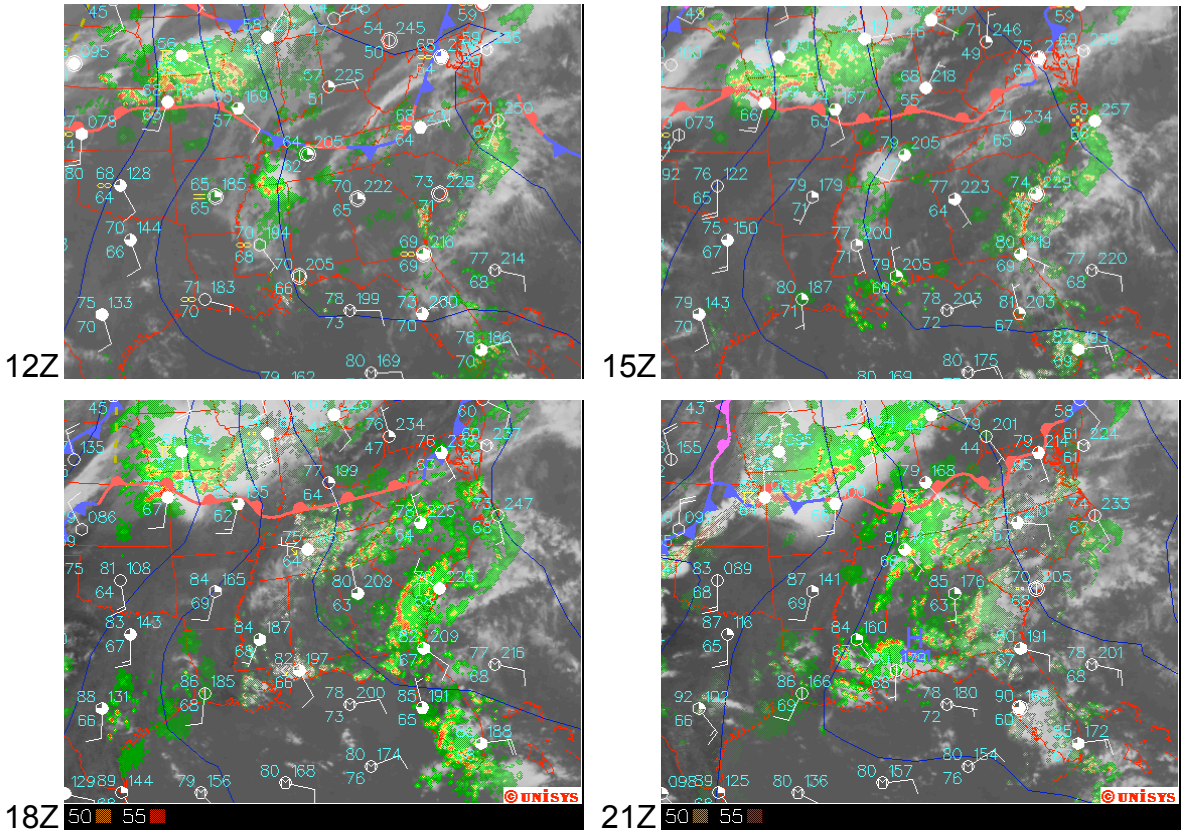


Fig. 17. Satellite images, surface observations and surface analysis on 15 May 2009.

The line system on the southeastern coast pushing inland seems to have been associated with a sea breeze front as well as other favorable mesoscale and synoptic conditions. The operational HRRR for the cycles at 12 Z and 15 Z on that day predicted some scattered convection on the coast and farther inland from 18 – 23 Z, but none of the observed line formation and movement (Fig. 18).

In the sensitivity test, the forecasts were rerun with the model grid moved farther toward the south and east to avoid lateral boundary interference. The reruns again failed to predict the line formation and movement (Fig. 18). Instead, they forecast scattered convection in a broad area similar to those in the operational HRRR. An examination of the model low-level wind field shows that there is no clearly defined convergence zone (signalizing a sea breeze front) in the forecasts at 18 – 22 Z. At the model initial time (12 Z or 15 Z), the low-level wind does show some convergence areas in the coastal region, but they are broad and segmented, and become increasingly noisy with model integration. As a result, the model predicted a broad area of unstable atmosphere, but it lacked a mechanism to organize the storms into a fine line, which is presumably a result of interactions between the storms and the sea breeze frontal environment, and understandably difficult to simulate.

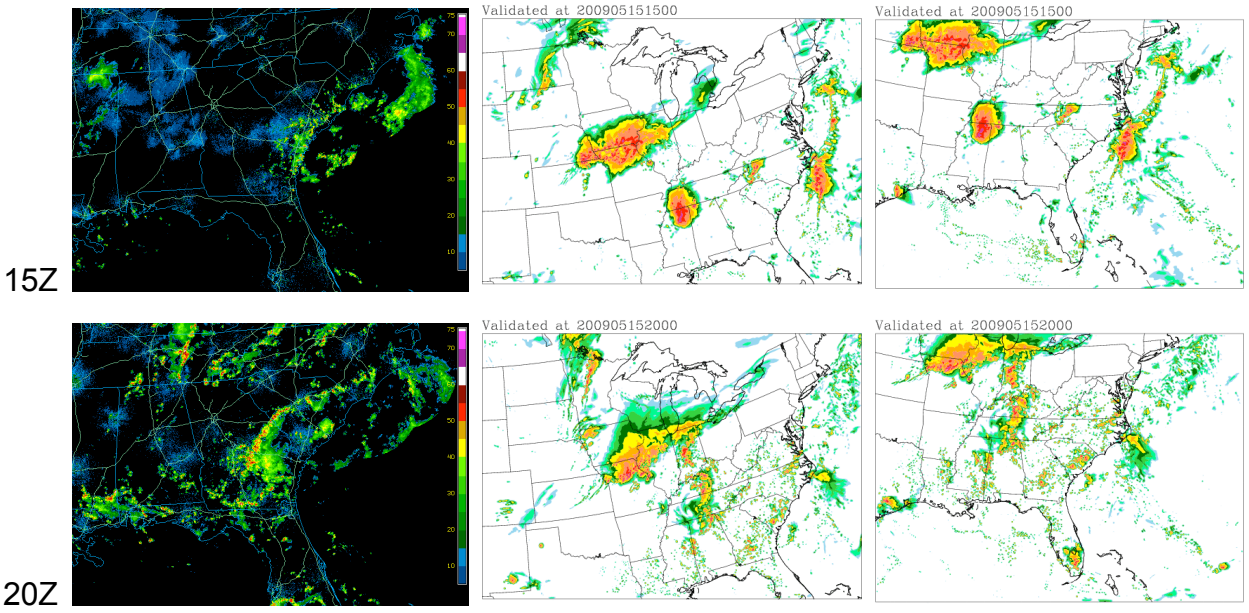


Fig. 18. Radar observation and model forecasts (of VIL) on 15 May 2009. Shown are 3 h and 8 h forecasts valid at 15 Z and 20 Z of 20090515. At the central column are operational HRRR forecasts and the right column are forecasts with domain moved southeast.

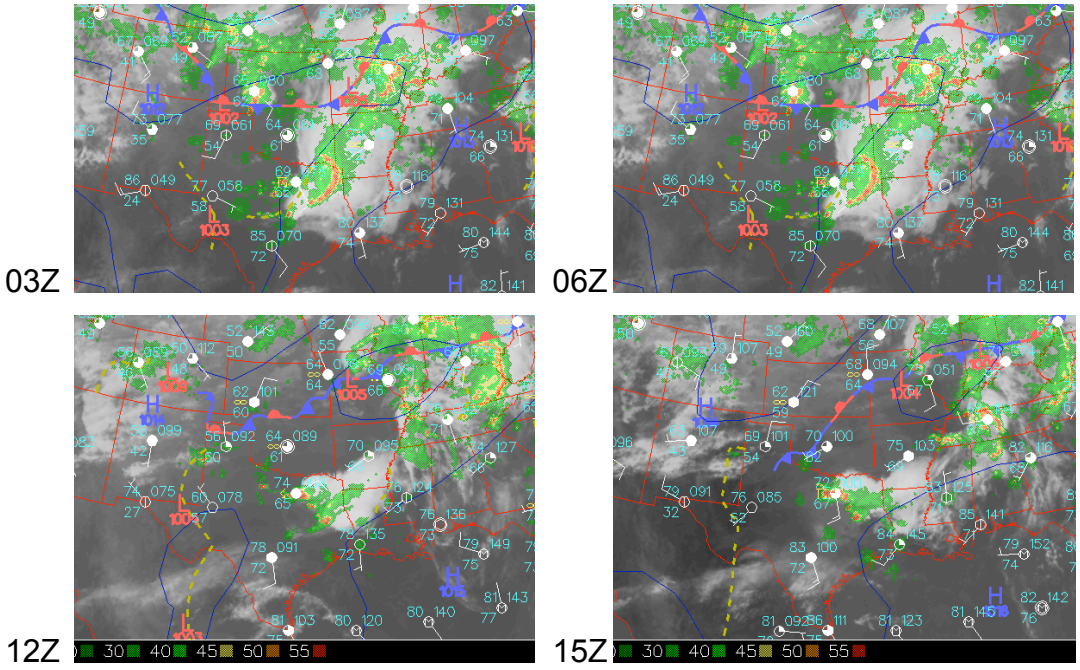


Fig. 19. Satellite images, surface observations and surface analysis on 11 June 2009.

On 11 June 2009, the system of interest started to form as a few cells of storms in Dallas-Fort Worth area around 3 Z, behind a boundary preceded by a major MCS

system (Fig. 19). As the storms moved eastward and southeastward, new cells continued to form in the original location and along a cold pool boundary, creating a train of storms and an elongated area of precipitation that lasted through 20 Z on 11 June 2009. The HRRR forecasts starting from 00 – 12 Z failed to forecast the system. In fact, the location of the system was too close to the HRRR southern boundary to allow a fair inspection.

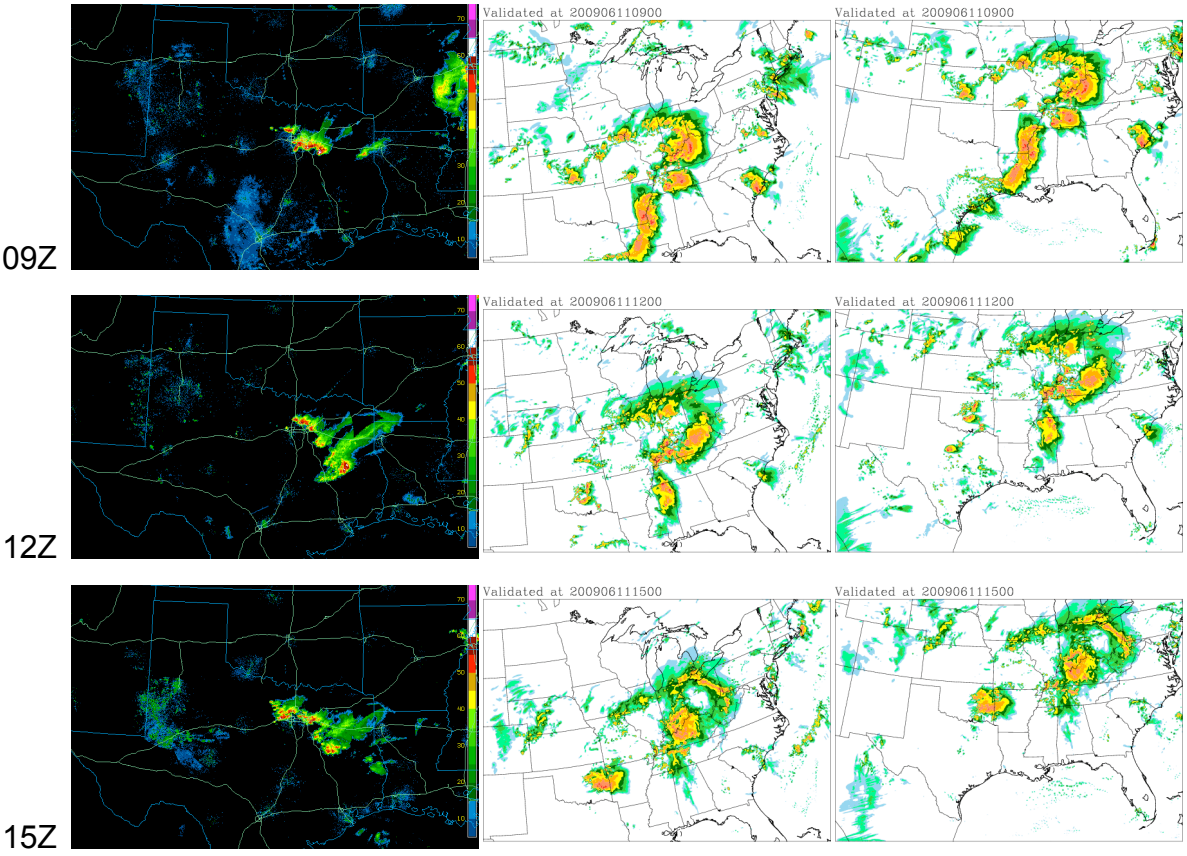


Fig. 20. Radar observation and model forecasts on 11 June 2009. Shown are 3 h, 6 h and 9 h forecasts valid at 09 Z, 12 Z and 15 Z of 20090611. At the central column are operational HRRR forecasts and the right column are forecasts with domain moved to the south.

The 00 Z and 06 Z forecasts of the day were re-run with the model southern boundary moved farther south by about 700 km. However, the rerun also failed to produce any significant, lasting system in the Dallas-Forth Worth area. The inaccurate depiction of the preceding MCS caused the model to place the cold pool boundary way too far south. The model instead predicted some scattered storms further south of the observed system (Fig. 20). Even though at the forecast initial time, say 06Z, the model utilized information from radar reflectivity, the details of the MCS, especially the cold pool boundary locations in the model, were not reproduced accurately. These case

studies suggest that the effect of lateral boundary proximity might be of secondary importance.

5. Summary and conclusions

The performance of the 3km HRRR was evaluated through a subjective (qualitative) assessment in real time by the SoD's and off-line (objective) sensitivity studies based on selected cases from the respective CoSPA domains used during the summers 2008 and 2009. The real-time assessment aimed at examining forecast success and failure of convective weather events, while the sensitivity studies using selected cases targeted the impact of initialization, grid increment, microphysics scheme, and lateral boundary effect on HRRR WRF-ARW short-term forecasts.

The qualitative assessment of the HRRR model based on both the CoSPA and VORTEX2 demonstrations suggests that the HRRR has a high probability of detection of severe storms but false alarms occur occasionally. The model seems to perform particularly well during the late afternoon and into nighttime, while the performance was not as good in the morning and early afternoon, with significantly less areal coverage of convection than was observed. The timing of HRRR forecasted and observed storms were typically within 1-2 hours of each other for storms that were of interest to VORTEX2. Two cases were examined to diagnose the sources of error that may have caused a poorer performance in the morning and early afternoon. The diagnosis based on surface observations indicates that the daytime low-level temperature contrasts between sunny and cloudy regions were typically less in the model than observed in the atmosphere. Analyses are ongoing to determine the generality of these results, and the implications for convective forecasts. Results of this ongoing work will be reported in the future.

The sensitivity studies using selected cases from the CoSPA 2008 and 2009 domains suggest that the initial conditions mattered most to the 0-12 hour forecasts of convective cases. The increase of grid resolution from 5km to 3km had greater impact than from 3km to 1km, although the latter seemed to show noticeable sensitivity beyond the 12 forecast range. The comparison between the current operationally-used Thompson microphysics scheme and the new two-moment Thompson scheme in WRFv3.1 showed only marginal increase in computation cost. However, it was found that the new scheme slightly reduced the model spin-up problem. Two severe convective cases that occurred near the model lateral boundary and that showed poor forecasts were examined to assess the effect of the lateral boundary setting. It was found that the lateral boundary proximity seems not to be the reason for the noted difficulty in successfully forecasting those severe storms. Although the results from these case studies provide useful knowledge about the characteristics and sensitivity of the HRRR forecasts, they cannot be generalized at this time. Additional case studies and in-depth analyses will have to be conducted to firm the preliminary results obtained so far.